CCE estimation of factor-augmented regression models with more factors than observables

Karabiyik, Hande; Urbain, Jean Pierre; Westerlund, Joakim

published in
Journal of Applied Econometrics
2019

DOI (link to publisher)
10.1002/jae.2661

document version
Publisher's PDF, also known as Version of record

Received: 16 October 2017 | Revised: 20 August 2018
DOI: 10.1002/jae.2661

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

CCE estimation of factor-augmented regression models with more factors than observables

Hande Karabiyik[1] | Jean-Pierre Urbain[2] | Joakim Westerlund[3,4]

[1]School of Business and Economics, Vrije Universiteit, Amsterdam, Netherlands

[2]Department of Quantitative Economics, Maastricht University, Maastricht, Netherlands

[3]Department of Economics, Lund University, Lund, Sweden

[4]Centre for Financial Econometrics, Deakin University, Geelong, Victoria, Australia

Correspondence
Joakim Westerlund, Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden.
Email: joakim.westerlund@nek.lu.se

Summary

This paper considers estimation of factor-augmented panel data regression models. One of the most popular approaches towards this end is the common correlated effects (CCE) estimator of Pesaran (Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 2006, *74*, 967–1012, 2006). For the pooled version of this estimator to be consistent, either the number of observables must be larger than the number of unobserved common factors, or the factor loadings must be distributed independently of each other. This is a problem in the typical application involving only a small number of regressors and/or correlated loadings. The current paper proposes a simple extension to the CCE procedure by which both requirements can be relaxed. The CCE approach is based on taking the cross-section average of the observables as an estimator of the common factors. The idea put forth in the current paper is to consider not only the average but also other cross-section combinations. Asymptotic properties of the resulting combination-augmented CCE (C^3E) estimator are provided and tested in small samples using both simulated and real data.

1 | INTRODUCTION

Consider the scalar and $m \times 1$ vector of observable panel data variables $y_{i,t}$ and $x_{i,t}$, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ index the cross-sectional and time series dimensions, respectively. The data-generating process (DGP) of the $T \times 1$ vector $y_i = [y_{i,1}, \ldots, y_{i,T}]'$ is similar to the DGP of Pesaran (2006) and is given by

$$y_i = x_i \beta_i + e_i, \tag{1}$$

$$e_i = F \lambda_i + \epsilon_i, \tag{2}$$

$$\beta_i = \beta + \xi_i, \tag{3}$$

where $x_i = [x_{i,1}, \ldots, x_{i,T}]'$ is $T \times m$, β_i is an $m \times 1$ vector of slope coefficients, $F = [F_1, \ldots, F_T]'$ is a $T \times r$ matrix of common factors with λ_i being the associated $r \times 1$ vector of factor loadings, $\epsilon_i = [\epsilon_{i,1}, \ldots, \epsilon_{i,T}]'$ is a $T \times 1$ vector of errors that are largely idiosyncratic, and ξ_i is an $m \times 1$ vector of errors.

The above model is the prototypical pooled panel regression with a factor error structure, in which ϵ_i is independent of x_i. If F is also independent of x_i, then Equation 1 is nothing but a static panel data regression with exogenous regressors, which can be estimated consistently using least squares (LS). If, however, x_i is correlated with F, then consistency will be

Our very good friend and co-author Jean-Pierre Urbain passed away shortly after the submission of the first version of this paper.

lost. To allow for this possibility, we follow Pesaran (2006) and assume that

$$x_i = F\Lambda_i' + \eta_i, \tag{4}$$

where Λ_i is an $m \times r$ loading matrix and $\eta_i = [\eta_{i,1}, \ldots, \eta_{i,T}]'$ is a $T \times m$ matrix of idiosyncratic errors. By combining Equations 1–4

$$w_i = FC_i + u_i, \tag{5}$$

where $w_i = [w_{i,1}, \ldots, w_{i,T}]'$ is $T \times (m+1)$, $w_{i,t} = [y_{i,t}, x_{i,t}']'$ is $(m+1) \times 1$, $C_i = [\Lambda_i' \beta_i + \lambda_i, \Lambda_i']$ is $r \times (m+1)$, and $u_i = [u_{i,1}, \ldots, u_{i,T}]' = [\eta_i \beta_i + \epsilon_i, \eta_i]$ is $T \times (m+1)$. Thus (1)–(4) can be rewritten equivalently as a static factor model for w_i, which is convenient because it means that the common component of the data can be estimated using existing methods for such models (see Chudik & Pesaran, 2015b, for a recent survey). In this paper, however, we focus on the CCE approach of Pesaran (2006), which has become very popular in the empirical literature, with a large number of applications.[1] The approach has also attracted much interest in the econometric literature, where it has been shown to work under very general conditions, including models with weak factors, dynamic models and even models with nonstationary data (see, e.g., Chudik et al., 2011; Chudik & Pesaran, 2015a; Kapetanios et al., 2011; Pesaran et al., 2013; Reese & Westerlund, 2016, 2018; Westerlund, 2018).

As is well known from the classical common factor literature, F and C_i are not separately identifiable, which means that the best that one can hope for is consistent estimation of the space spanned by F. The idea of Pesaran (2006) was to make use of the cross-section variation to estimate this space. A natural way to accomplish this is to take the cross-section average, giving $\overline{w}_t = \overline{C}' F_t + \overline{u}_t$, where $\overline{C}$, $\overline{w}_t$ and $\overline{u}_t$ are the cross-section averages of C_i, $w_{i,t}$ and $u_{i,t}$, respectively. Hence, since $\overline{u}_t \to_p 0_{(m+1)\times 1}$ as $N \to \infty$, where $\to_p$ signifies convergence in probability, we have that $\overline{w}_t = \overline{C}' F_t + \overline{u}_t \to_p \overline{C}' F_t$. This suggests using $\overline{w}_t$ as an estimator of $\overline{C}' F_t$, a strategy that would seem to require

$$\operatorname{rk} \overline{C}' = r \leq m+1, \tag{6}$$

where rk$A$ denotes the rank of any matrix A. Hence the number of observables must be at least as large as the number of factors. The idea behind the CCE approach is then to estimate β from a pooled LS regression of $y_{i,t}$ onto $x_{i,t}$ and $\overline{w}_t$, leading to the pooled CCE (CCEP) estimator.[2]

Interestingly, as Pesaran (2006) pointed out, the condition in Equation 6 is actually not necessary when using the CCEP estimator. However, as has been shown by Westerlund and Urbain (2013), and as we explain in detail in Section 3 of the current paper, relaxing Equation 6 requires imposing additional restrictive independence conditions on λ_i and Λ_i, which, if false, may well render the CCEP estimator inconsistent. Hence, even if Equation 6 can in principle be relaxed, in most situations of practical relevance this is not necessarily so. Also, even if the more restrictive assumptions are satisfied, the rate of consistency of the CCEP estimator when $r > m+1$ is lower than when $r \leq m+1$.

The discussion in the previous paragraph suggests that it is important to have $r \leq m+1$. The question therefore arises as to how likely this is in practice. The number of regressors, m, is usually a small number that is given by economic theory (and/or previous empirical evidence). On the other hand, economic theory is not very informative regarding the number of factors r (see, e.g., Eberhardt et al., 2013). Therefore, the theoretically implied value of m typically has little or nothing to do with r. This is important because within CCE choosing m also means restricting r, and in many applications there is little or no reason to believe that this number should be less than or equal to $m+1$. For example, Su et al. (2015) regress the gross domestic product (GDP) growth of a country on its investment share.[3] They applied the information criteria (IC) of Bai and Ng (2002), with which they found evidence of between three and four factors. Another interesting example is provided by Holly et al. (2010), who used CCE to study the effect of income on housing prices in the USA. They applied the same IC as Su et al., and reported evidence of up to six factors. Hence, in both examples, r is estimated to be larger than $m+1$. There are many more examples like these.[4] In view of this and the potential problems involved when

[1]In the supplementary material, provided online as Supporting Information, we present some Monte Carlo results that enable comparison with the principal components-based estimator considered by Greenaway-McGrevy et al. (2012), Westerlund and Urbain (2015), and Chudik et al. (2011), which is arguably one of the closest competitors of the CCE approach.

[2]Another possibility is to estimate β_i from a time series LS regression of $y_{i,t}$ onto $x_{i,t}$ and $\overline{w}_t$. This is the individual-specific CCE estimator, which can be averaged across the cross-section to obtain the mean group CCE (CCEMG) estimator. However, for reasons to be explained in Section 3, in this paper we focus on the CCEP estimator, although we also discuss the results for the other CCE estimators.

[3]Su et al. (2015) also included the country's initial economic condition as a regressor. However, since this variable is time invariant it cannot be used in the factor estimation stage.

[4]Most evidence on the number of common factors is based on univariate models. A prominent example is returns, where multifactor models such as the three-factor model of Fama and French (1996) have attracted considerable attention. In fact, there is by now quite some evidence suggesting that

$r > m + 1$, the restriction in Equation 6 cannot be given but should really be tested on a case-to-case basis. In practice, however, this aspect is almost always ignored.

In the current paper we take this shortcoming as our starting point. The purpose is to provide a simple modification of the original CCE approach allowing (but not requiring) $r > m + 1$. Hence the purpose here is not really to propose an alternative estimator, but to show that original CCE belongs to a much broader class of estimators, which is henceforth referred to as combination-augmented CCE (C^{3}E). The idea behind C^{3}E is to consider not only the equal-weighted cross-section average but also other combinations of $w_{1,t}, \ldots, w_{N,t}$. In particular, by considering k such combinations, we can allow for

$$k(m+1) \geq m+1$$

common factors. In addition to the larger number of factors that can be allowed, the new approach also enables one to consider separately the selection of m and r, which is again not possible within the original CCE framework.

The remainder of the paper is organized as follows. Section 2 lays out the assumptions under which we will be working. Some of the these are more restrictive than necessary, but are kept for ease of exposition. In the supplementary Supporting Information we provide a set of more relaxed conditions that are used to establish the asymptotic distribution of the pooled C^{3}E estimator. This is done in Section 3. We show that the estimator is consistent and asymptotically normal with the rate of consistency depending on whether the slopes are homogeneous or not. The results are established under the condition that $N, T \to \infty$ with $T/N \to \tau < \infty$, although in the paper we focus on the case when $T/N \to 0$. As a solution to the practical problem of how to pick the appropriate combinations, an IC-based selection rule is proposed in Section 4. Section 5 presents the results of an empirical application using as an example the gravity equation of trade. Section 6 concludes. All proofs are provided in the supplementary Supporting Information, which also reports the results of a large-scale Monte Carlo study.

2 | ASSUMPTIONS

In the supplementary Supporting Information, we provide a set of minimal conditions on ϵ_i, η_i, F, β_i, λ_i, and Λ_i under which our results hold. These are so-called "high-level" assumptions, and are similar to those used by Bai and Ng (2002) and Bai (2003, 2009), for example. The advantage of making such high-level assumptions is that the results cover a wide range of DGPs. The obvious disadvantage is that the assumptions are difficult to interpret, and in this section we therefore provide a set of more primitive, easy-to-interpret conditions that satisfy the minimal ones. Here and throughout this paper, $\operatorname{tr} A$, and $\|A\| = \sqrt{\operatorname{tr}(A'A)}$ will be used to denote the trace and Frobenius (Euclidean) norm, respectively, of the matrix A. Also, $\to_d$ and $\to_p$ signify convergence in distribution and probability, respectively.

Assumption 1.

(i) $\epsilon_{i,t}$ and $\eta_{i,t}$ are independent and identically distributed (i.i.d.) across i, and follow stationary linear processes with zero means, $E(\epsilon_{i,t}^2) = \sigma_{\epsilon,i}^2$, $E(\eta_{i,t}\eta_{i,t}') = \Sigma_{\eta,i}$, and absolutely summable autocovariances. Also, $N^{-1}\sum_{i=1}^{N}\Sigma_{\eta,i} \to \Sigma_\eta$ as $N \to \infty$, where Σ_η is positive definite.

(ii) F_t is covariance stationary such that $E(\|F_t\|^4) < \infty$ and $T^{-1}\sum_{t=1}^{T} F_t F_t' \to_p \Sigma_F$ as $T \to \infty$, where Σ_F is positive definite.

(iii) ξ_i is i.i.d. across i with $E(\xi_i) = 0_{m\times 1}$, $E(\xi_i\xi_i') = \Sigma_\xi$ positive definite, and $E(\|\xi_i\|^4) < \infty$.

(iv) λ_i and Λ_i are nonrandom such that $\|\lambda_i\| < \infty$ and $\|\Lambda_i\| < \infty$.

(v) $\epsilon_{i,t}$, $\eta_{j,s}$, F_l, and ξ_n are mutually independent for all i, t, j, s, l and n.

Remark 1. Assumptions 1(i)–(iii) and (v) are the same as in Pesaran (2006), and we therefore refer to that paper for a thorough discussion. Here we only comment on the most important assumptions. Assumption 1(i) allows for weak serial correlation in the idiosyncratic errors, but not unit roots, and is standard in the literature. Similarly, while F_t can be serially correlated, it cannot contain unit roots or any other trends for that matter, as $T^{-1}\sum_{t=1}^{T} F_t F_t'$ would not converge in this case. As with (i), condition (ii) is standard in the literature, and we will keep it here for comparability. However, we would like to point out that both assumptions can actually be relaxed. In the supplementary Supporting

three factors are not enough, and that proper accounting for the observed cross-firm correlation might potentially require a large number of factors (see, e.g., Bai & Ng, 2006b). Similar evidence has been found also for other variables (see, e.g., Greenaway-McGrevy et al., 2012; Moon & Weidner, 2015).

Information, we show how (i) can be relaxed to allow for weak error cross-sectional correlation without affecting the results. Condition (ii) can be relaxed along the lines of Westerlund (2018) to allow for more general types of factors (see also Kapetanios et al., 2011), and in the supplementary Supporting Information we report some confirmatory Monte Carlo results to this effect. In Section 5, we demonstrate the usefulness of being able to relax Assumption 1(ii) in practice, and discuss the type of factors that can be permitted.

The main difference when compared to Pesaran (2006) is Assumption 1(iv). Specifically, while in Pesaran (2006, Assumption 3) λ_i and Λ_i are assumed to be i.i.d. and also independent of each other, under our Assumption 1(iv) λ_i and Λ_i are treated as fixed parameters, which is a more general consideration. Note in particular how the correlation between λ_i and Λ_i is not restricted in any way. In Section 3, we elaborate on this.

For each of the $m + 1$ columns in w_i, we consider k cross-section combinations, as given by the $T \times k(m + 1)$ matrix $N^{-1}\sum_{i=1}^{N} w_i Z_i$, where $Z_i = (I_{m+1} \otimes z_i')$ is $(m + 1) \times k(m + 1)$ and $z_i = [z_{1,i}, \dots, z_{k,i}]'$ is a $k \times 1$ vector of combinations. The combinations can be deterministic and/or stochastic, provided that Assumption 2 is satisfied. Here and throughout this paper

$$\overline{H} = \frac{1}{N}\sum_{i=1}^{N} Z_i' C_i', \tag{7}$$

a $k(m + 1) \times r$ matrix.

Assumption 2.

(i) $\mathrm{rk}\overline{H} = k(m+1) = r$ for all $N < \infty$ and $\overline{H} \to_p H$ as $N \to \infty$, where $\mathrm{rk}\, H = k(m + 1) = r$ and $\|H\| < \infty$.
(ii) Z_i is either deterministic such that $\|Z_i\| < \infty$, or stochastic such that $E(\|Z_i\|^2) < \infty$. In the latter case, we also assume that Z_i is uncorrelated with $u_{j,t}$ for all i, j and t.

Remark 2. Note that if $k = 1$ and $z_i = 1$, then $N^{-1}\sum_{i=1}^{N} w_i Z_i = N^{-1}\sum_{i=1}^{N}(w_i \otimes z_i') = \bar{w}$, and so we are back in the cross-section average-only original CCE approach of Pesaran (2006). In this case, Assumption 2 is the same as in Pesaran, in the sense that (i) boils down to Equation 6 with $r = m + 1$, and (ii) is trivially satisfied. Pesaran does point out that the equal-weighted average is not the only way to combine the data. However, while recognizing the fact that the weights do not have to be equal, it is still just one combination/weighted average per observable that is being considered. The contribution of the present paper is the consideration of multiple combinations, which is important, because it relaxes the $m + 1 \geq r$ requirement in Equation 6. This makes it necessary to be specific about the combinations that can be permitted. Interestingly, z_i can be thought of as acting as an instrument for C_i. Assumption 2 is therefore analogous to the well-known orthogonality and validity conditions in the instrumental variables (IV) literature (see Bai & Ng, 2010, for a panel IV approach based on similar assumptions). Specifically, z_i should be orthogonal in the sense that it should be uncorrelated with $u_{i,t}$, and it should be valid in the sense that $\mathrm{rk}\bar{H} = k(m+1) = r$. The combinations should therefore be correlated with the loadings, and the loadings measure the extent to which the cross-sectional units are affected by the common factors. The combinations should therefore be informative about the effect of the factors. In Section 4 we elaborate on this. We also propose an IC-based procedure that selects only the valid combinations from a set of possible candidates. In this sense, the assumption that $k(m + 1) = r$, as apposed to $k(m + 1) \leq r$, is without loss of generality, which is a standard argument in the part of the factor-augmented regression literature that is based on estimated principal components (PC) factors (see, e.g., Bai, 2009; Greenaway-McGrevy et al., 2012). Another reason for assuming $k(m + 1) = r$ is that it simplifies the interpretation of the outcome of the IC, as we explain in detail in Section 4.

Finally, note that the factors cannot be weak, as this would make $\overline{H}$ rank deficient in the limit as $N \to \infty$. The extension to the weak factor case should, however, be relatively straightforward following the steps of Chudik et al. (2011) in the CCE context.

3 | C^{3}E ESTIMATION AND INFERENCE

In Section 3.1 we study the asymptotic properties of the pooled C^{3}E estimator in the case when $\beta_1, \dots, \beta_N$ are all equal, and in the case when they are unrestricted. The estimation of the various covariance matrices that appear in Section 3.1 is discussed in Section 3.2.

3.1 | The pooled C^{3}E estimator

As already mentioned, since F and C_i are not separately identifiable, F can only be estimated up to a matrix rotation. The proposed estimator $\hat{F}$ of $F\overline{H}$ is given by

$$\hat{F} = \frac{1}{N}\sum_{i=1}^{N} w_i Z_i = \frac{1}{N}\sum_{i=1}^{N}\left(w_i \otimes z_i'\right), \tag{8}$$

whose dimension is $T \times k(m+1)$. The resulting pooled estimator of β is given by

$$\hat{\beta}_P = \left(\sum_{i=1}^{N} x_i' M_{\hat{F}} x_i\right)^{-1} \sum_{i=1}^{N} x_i' M_{\hat{F}} y_i, \tag{9}$$

where $M_A = I_T - A(A'A)^{-1}A'$ for any T-rowed matrix A. The CCEP estimator, henceforth denoted by $\hat{\beta}_{\mathrm{CCEP}}$, is simply $\hat{\beta}_P$ with $\hat{F} = \bar{w}$.

Remark 3. The pooled C^{3}E estimator considered here is based on "within" pooling, whereby the data are summed over the cross-section before taking the ratio. Another approach is to use "between" pooling, in which case the ratio is taken prior to summing over the cross-section. Pesaran (2006) considers both types of pooling. However, since in his Monte Carlo study within pooling generally leads to the best performing estimator, in this paper we only consider this type. However, as mentioned above, as a by-product of the need for consistent covariance estimation in the heterogeneous slope case, in Section 3.2 we also consider the individual-specific C^{3}E estimator. This estimator can be averaged, leading to a between (or "group mean") type C^{3}E estimator.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied with* $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$. *As* $N, T \to \infty$ *with* $T/N \to 0$,

$$\sqrt{NT}(\hat{\beta}_P - \beta) \to_d N\left(0_{m\times 1}, \Sigma_\eta^{-1} S \Sigma_\eta^{-1}\right),$$

where

$$S = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} S_i,$$
$$S_i = \lim_{T\to\infty} T^{-1} E(\eta_i' \Omega_{\epsilon,i} \eta_i'),$$

with $\Omega_{\epsilon,i} = E(\epsilon_i \epsilon_i')$.

Theorem 1 is concerned with the conventional homogeneous slope case, and is the C^{3}E counterpart of Theorem 4 of Pesaran (2006), which requires that $r = 1$. Theorem 1 only requires that $k(m+1) = r$ and is therefore more general in this regard. The theorem supposes that $T/N \to 0$, which is the same condition as in Theorem 4 of Pesaran. The proof provided in the supplementary Supporting Information is, however, more general in that it only requires $T/N \to \tau < \infty$. The results show that the asymptotic distribution of $\sqrt{NT}(\hat{\beta}_P - \beta)$ is not correctly centered at zero when $\tau > 0$. According to our Monte Carlo results, however, the C^{3}E estimator tends to perform very well in small samples, even in cases when $T > N$. We therefore do not consider the $\tau > 0$ case here, but put it in the supplementary material. The supplement also introduces a bias-corrected C^{3}E estimator and establishes its validity.

Analogous to the bulk of the existing literature on factor-augmented regressions (see, e.g., Bai, 2009; Gonçalves & Perron, 2014; Greenaway-McGrevy et al., 2012; Moon & Weidner, 2015), Theorem 1 supposes that $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$. The effect of a violation of this assumption is studied in Theorem 2.

Theorem 2. *Under Assumptions 1 and 2, as* $N, T \to \infty$ *with* $\sqrt{N}/T \to 0$,

$$\sqrt{N}(\hat{\beta}_P - \beta) \to_d N\left(0_{m\times 1}, \Sigma_\eta^{-1} R \Sigma_\eta^{-1}\right),$$

where

$$R = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \Sigma_{\eta,i} \Sigma_\xi \Sigma_{\eta,i}.$$

Theorem 2 is the C^{3}E counterpart of Theorem 3 of Pesaran (2006). It shows that the variance of the estimator emanates from the heterogeneity of the slopes, as measured by Σ_ξ. This result is analogous to that of Pesaran for the CCEP estimator. However, the asymptotic variance of this estimator has an additional term that depends on the heterogeneity of the factor loadings and that is there because the rank condition in Equation 6 is not assumed to be met. The C^{3}E estimator does not depend on whether Equation 6 is satisfied, which is also the reason why the asymptotic distribution given in Theorem 2 does not depend on the factor loadings. In order to illustrate this point, suppose first that $m+1=r$. Since in this case $\overline{w}_t = \overline{H}' F_t + o_p(1)$, where $\overline{H} = \overline{C}$ is of full rank and hence invertible, we have $N^{-1/2}T^{-1}\sum_{i=1}^N x_i' M_{\overline{w}} F\lambda_i = N^{-1/2}T^{-1}\sum_{i=1}^N x_i' M_{F\overline{H}} F\lambda_i + o_p(1) = N^{-1/2}T^{-1}\sum_{i=1}^N x_i' M_F F\lambda_i + o_p(1) = o_p(1)$ (see Pesaran, 2006, Equation 40). Hence

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{\mathrm{CCEP}} - \beta) &= \left(\frac{1}{NT}\sum_{i=1}^N x_i' M_{\overline{w}} x_i\right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^N x_i' M_{\overline{w}}(x_i\xi_i + F\lambda_i + \epsilon_i) \\ &= \left(\frac{1}{NT}\sum_{i=1}^N x_i' M_{\overline{w}} x_i\right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^N x_i' M_{\overline{w}} x_i\xi_i + o_p(1),\end{aligned} \tag{10}$$

which converges to the same asymptotic distribution given in Theorem 2, provided that ξ_i is "nicely" behaved in the sense that Assumption 1(iii) is met. If, on the other hand, $m+1<r$, then $N^{-1/2}T^{-1}\sum_{i=1}^N x_i' M_{\overline{w}} F\lambda_i$ will not be negligible (see Pesaran, 2006, Equation 38), and so we obtain

$$\sqrt{N}(\hat{\beta}_{\mathrm{CCEP}} - \beta) = \left(\frac{1}{NT}\sum_{i=1}^N x_i' M_{\overline{w}} x_i\right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^N x_i' M_{\overline{w}}(x_i\xi_i + F\lambda_i) + o_p(1), \tag{11}$$

which will not converge in distribution unless λ_i is also nicely behaved. As pointed out by Westerlund and Urbain (2013), one possibility here is to assume that λ_i and Λ_i are mutually independent, which seems like a rather restrictive assumption. For example, when regressing investments on savings, as is commonly done in the literature on the so-called "Feldstein–Horioka puzzle," a common shock that increases savings is going to push interest rates down and investments up, suggesting that λ_i and Λ_i should be negatively correlated. Thus, while the requirement that $m+1 \geq r$ can be relaxed also within the original CCE framework, this does not come free of charge.

It is important to note that in the above example the rate of consistency of $\hat{\beta}_{\mathrm{CCEP}}$ is given by $\sqrt{N}$ and not by $\sqrt{NT}$. One may think that this relatively low rate of consistency is due to the heterogeneity of β_i, and that imposing $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$ would prevent this from happening, regardless of whether $m+1 \geq r$ or $m+1<r$.[5] However, this is not the case. The reason is easily appreciated by simply imposing $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$ and using $(NT)^{-1/2}\sum_{i=1}^N x_i' M_{\overline{w}}\epsilon_i = O_p(1)$, from which it follows that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N x_i' M_{\overline{w}}(x_i\xi_i + F\lambda_i + \epsilon_i) = \frac{1}{\sqrt{NT}}\sum_{i=1}^N x_i' M_{\overline{w}} F\lambda_i + O_p(1). \tag{12}$$

If $m+1 \geq r$, then $(NT)^{-1/2}\sum_{i=1}^N x_i' M_{\overline{w}} F\lambda_i = o_p(1)$, and so we obtain $\sqrt{NT}(\hat{\beta}_{\mathrm{CCEP}} - \beta) = O_p(1)$. Hence, provided that $m+1 \geq r$, imposing $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$ restores $\sqrt{NT}$-consistency. If, on the other hand, $m+1<r$, then $T^{-1}x_i' M_{\overline{w}} F = O_p(1)$, and therefore

$$\sqrt{NT}(\hat{\beta}_{\mathrm{CCEP}} - \beta) = \sqrt{T}\left(\frac{1}{NT}\sum_{i=1}^N x_i' M_{\overline{w}} x_i\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^N T^{-1} x_i' M_{\overline{w}} F\lambda_i + O_p(1), \tag{13}$$

whose order is determined by the order of the first term on the right, which in turn depends on λ_i and Λ_i. If λ_i is i.i.d. and independent of Λ_i, then the first term is $O_p(\sqrt{T})$, whereas if λ_i is non-i.i.d. and/or correlated with Λ_i, then the same term is $O_p(\sqrt{NT})$. Thus the rate of consistency is $\sqrt{N}$, at best, and if λ_i is non-i.i.d. and/or correlated with Λ_i, then $\hat{\beta}_{\mathrm{CCEP}}$ is even inconsistent. The proposed C^3E estimator in the homogeneous slope case is not only very simple but also $\sqrt{NT}$-consistent regardless of the specification of λ_i and Λ_i, provided that Assumptions 1 and 2 are satisfied.

In the supplementary Supporting Information, we use Monte Carlo simulations to evaluate the small-sample performance of the CCE and C^3E estimators. The results are largely in agreement with our expectations given the above discussion. In particular, we find that whenever Equation 6 is satisfied the performance of the CCE and C^3E estimators are comparable. If, however, Equation 6 is not met, then the C^3E estimator continues to work well, while the CCE estimator breaks down.

[5]It is not clear from Pesaran (2006) whether one can have $\xi_1 = \cdots = \xi_N = 0_{m\times 1}$, while at the same time permitting $m+1<r$.

3.2 | Covariance matrix estimation

In this section, we derive consistent estimators of the covariance matrices that appear in Theorems 1 and 2. We begin by considering $\Sigma_\eta^{-1}S\Sigma_\eta^{-1}$, which according to Theorem 1 is the appropriate covariance matrix to consider when $\beta_1, \ldots, \beta_N$ are all equal. Let $\hat{e}_i = [\hat{e}_{i,1}, \cdots, \hat{e}_{i,T}]' = M_{\hat{F}}(y_i - x_i\hat{\beta}_P)$ and $\hat{\eta}_i = [\hat{\eta}_{i,1}, \cdots, \hat{\eta}_{i,T}]' = M_{\hat{F}}x_i$. A naturally consistent estimator of $\Sigma_{\eta,i}$ is given by

$$\hat{\Sigma}_{\eta,i} = \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_{i,t}\hat{\eta}_{i,t}', \tag{14}$$

from which we obtain

$$\hat{\Sigma}_\eta = \frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{\eta,i}. \tag{15}$$

For $\Sigma_{\eta\epsilon}$, we follow Pesaran (2006), who recommend using a heteroskedasticity and autocorrelation consistent (HAC) estimator in the spirit of Newey and West (1987). The particular estimator considered here is given by

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N}\hat{S}_i, \tag{16}$$

where

$$\hat{S}_i = \hat{S}_i(0) + \sum_{j=1}^{p}\left(1 - \frac{j}{p+1}\right)[\hat{S}_i(j) + \hat{S}_i(j)'], \tag{17}$$

$$\hat{S}_i(j) = \frac{1}{T}\sum_{t=j+1}^{p}\hat{e}_{i,t}\hat{e}_{i,t-j}\hat{\eta}_{i,t}\hat{\eta}_{i,t-j}', \tag{18}$$

with p being the window size. The appropriate covariance matrix estimator to use in the homogeneous slope case is given by $\hat{\Sigma}_\eta^{-1}\hat{S}\hat{\Sigma}_\eta^{-1}$.

If, as in Theorem 2, $\beta_1, \ldots, \beta_N$ are not all the same, the above covariance estimator is no longer consistent. Specifically, while $\hat{\Sigma}_\eta$ is still consistent, because of the reduced rate of consistency of $\hat{\beta}_P$, $\hat{S}$ is inconsistent for R. Recognizing this problem, Pesaran (2006) proposed a nonparametric method that made use of the individual-specific CCE estimator. The appropriate C^3E analog of this estimator is given by

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N}\hat{R}_i, \tag{19}$$

where

$$\hat{R}_i = \hat{\Sigma}_{\eta,i}\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)'\hat{\Sigma}_{\eta,i}, \tag{20}$$

$$\hat{\beta}_i = (x_i'M_{\hat{F}}x_i)^{-1}x_i'M_{\hat{F}}y_i. \tag{21}$$

The consistency of this estimator follows from the consistency of the individual-specific C^3E estimator, $\hat{\beta}_i$, which is established in the supplementary material. The fact that $\hat{\beta}_i$ is consistent means that the appropriate covariance matrix estimator to consider in the heterogeneous slope case is given by $\hat{\Sigma}_\eta^{-1}\hat{R}\hat{\Sigma}_\eta^{-1}$.

4 | SELECTING THE COMBINATIONS

A problem in applications is how to construct the combination matrix, z. This problem can be seen as being comprised of two parts: (i) finding candidate combinations; and (ii) selecting among the candidates.

4.1 | Finding combination candidates

While z_i is not required to be uncorrelated with $u_{i,t}$, the correlation is not irrelevant, as the rate of consistency of $\hat{F}$ is increased when z_i and $u_{i,t}$ are uncorrelated. The combinations in z_i are therefore ideally chosen to be uncorrelated with $u_{i,t}$. They should also be highly correlated with C_i. Specifically, since $C_i = [\Lambda i'\beta_i + \lambda_i, \Lambda i']$, for z_i to be highly correlated with C_i, one should choose combinations that are believed to be highly correlated with the factor loadings.

An obvious approach to selecting the combinations is to exploit whether there are natural candidates in the particular application being considered, and we would like to argue that this scenario is in fact quite common. Indeed, the studies that express a suspicion of cross-section dependence without providing any explanation for why are few. For example, in macroeconomics usual common factor suspects include trade of goods and services, technology spillovers, and worldwide supply shocks, such as oil price shocks (see, e.g., Dees et al., 2007). The task of finding combinations that are correlated with the loadings is therefore tantamount to finding variables that measure the extent to which countries are affected by these usual suspects.

Mastromarco et al. (2016), and Eberhardt and Teal (2011) argue that spillover effects of globalization and business cycles, and common political, economic and spatial stimuli are likely to make production correlated across countries. As examples of variables that measure the effect of these common factors they mention openness, trade agreements, physical and human capital, growth determinants, initial per capita income, institutional environment, qualitative features of governance, geographical features, adoption of efficiency-enhancing technology, and natural resource constraints. If the analysis is made at the firm level, the extent to which a firm's production function is affected by common factors is likely to depend on the size of the firm, financial constraints, and the technology adopted, for example (see Chudik & Straub, 2011; Eberhardt & Teal, 2011). In the spillover literature, absorptive capacity is known to be an important determinant of the effect of knowledge, which can in turn be measured using, for example, openness, trade flows, human capital, and various development indices (see, e.g., Fracasso & Marzetti, 2014, 2015).

Baxter and Kouparitsas (2004) and Imbs (2004) studied the determinants of business cycle comovements. They concluded that trade was the most important determinant of cross-country business cycle linkages. Trade is an important determinant of cross-country linkages also in financial markets (see, e.g., Dees et al., 2007; Forbes & Chinn, 2004), although when modeling returns it is standard practice to use asset-specific characteristics ("fundamentals") such as industry classification, market capitalization, and style classification as observable loadings, or "betas" (Rosenberg, 1974).

A special kind of dependence that often arises is if there is a dominant unit (see Pesaran & Chudik, 2013). For example, in multicountry studies, it is natural to think of the USA as a dominant unit, whose influence may be measured using US trade weights and financial linkages, such as US equity and debt weights (see, e.g., Pesaran et al., 2004). Another example, taken from Holly et al. (2011), is the modeling of house prices in different regions in the UK, where price movements in London have had a dominating effect on all other regions.

Many of the above-mentioned sources of correlation are examples of what Conley (1999) referred to as "economic distance," and there are many more examples like these (see, e.g., Conley, 1999; Conley & Topa, 2002). Another important source of correlation is physical distance. Indeed, as Tobler (1970) points out in his "first law of geography," "everything is related to everything else; but near things are more related than distant things." A recent example of this is provided by Holly et al. (2010), who studied the effect of income on house prices across US states, and reported evidence suggesting that bordering states tend to be more correlated than states that are located further apart. Another example is provided in the empirical illustration reported in Section 5, where we consider a panel of bilateral trade flows among 15 European countries. Previous studies have shown that distance is important for trade, and our results confirm this.

As the above discussion illustrates, in many applications there are natural combination candidates that can go into z_i. Deterministic combinations are particularly simple to come by. Specifically, as Chudik et al. (2011) showed, the cross-section average can be quite effective in mopping up cross-section dependence. A vector of ones is therefore a good starting point. Tutz and Binder (2007) considered the problem of boosting ridge regression. They separated between "must have" candidates and other variables. The cross-section average can therefore be thought of as a "must have combination." This special treatment of the cross-section average highlights the role of C^{3}E as an extension rather than as an alternative to original CCE. Other readily available combination candidates are preliminary consistent estimates of (the space spanned by) C_i. The only requirement is that the rate of consistency must be at least $\sqrt{N}$, which is sufficiently relaxed to enable estimation by PC (see Bai, 2003).

4.2 | An IC-based selection procedure

An advantage of using deterministic combinations and/or preliminary loading estimates is that they are (asymptotically) uncorrelated with $u_{i,t}$.[6] However, in practice there is no guarantee that the combinations are valid, and if some of the combinations are stochastic there is also likely to be uncertainty regarding the correlation with $u_{i,t}$. In this section, we

[6]This advantage of using deterministic instruments has been pointed out before by Phillips and Hansen (1990) in the context of IV estimation of cointegrated time series regressions.

propose a selection criterion for the combination vector, z_i. In so doing, it is convenient, albeit not necessary (as we discuss at the end of this section) to assume that the valid combinations, henceforth denoted by $z_{0,i}$, are ordered first (for similar assumptions (see, e.g., Donald & Newey, 2001; Zheng & Loh, 1995, 1997), and that their number is given by k. In fact, analogous to IV selection, it is useful to treat also the combinations within $z_{0,i}$ as ordered, but then according to their correlation with C_i. The first combination in $z_{0,i}$ has the highest correlation. The invalid, or "nuisance," combinations, henceforth denoted by $z_{1,i}$, are ordered least, implying that z_i can be partitioned as $z_i = [z_{0,i}', z_{1,i}']'$, a $k_{\max} \times 1$ vector with $k_{\max} \geq k$.

Remark 4. As usual, the correct model must be among the models considered in the selection or else the selection procedure will be inconsistent. In our case, this is tantamount to requiring that $k_{\max} \geq k$, which is less restrictive than in many other model selection problems. The reason is that the number of allowable factors does not increase by a factor of 1 for each additional combination but rather by a factor of $m + 1$.

The IC considered in the present paper can be seen as a multivariate version of the IC criterion of Bai and Ng (2002) and is given by

$$\text{IC}(s) = \ln\det V(\hat{F}(s)) + s \cdot g, \tag{22}$$

where $V(A) = (NT)^{-1}\sum_{i=1}^{N} w_i' M_A w_i$ for any T-rowed matrix A, $\hat{F}(s)$ is $\hat{F}$ based on s combinations, and g is a penalty term. The associated IC estimator $\hat{k}$ of k is given simply by

$$\hat{k} = \underset{s=1,\ldots,k_{\max}}{\arg\min}\, \text{IC}(s). \tag{23}$$

Theorem 3. *Suppose that Assumptions 1 and 2 are met with z_i replaced by $z_{0,i}$, and that $g \to 0$ and $\min\{N, \sqrt{T}\} \cdot g \to \infty$, as $N, T \to \infty$. Then*

$$P(\hat{k} = k) \to 1.$$

Bai and Ng (2002) propose several ICs for selecting r in the context of PC estimation of common factor models. The Theorem 3 requirement that $g \to 0$ and $\min\{N, \sqrt{T}\} \cdot g \to \infty$ is stronger than in their paper. The reason for this difference is the invalid combination candidates, which are not restricted in any way and may therefore be correlated with $u_{i,t}$. The relatively large penalty employed here is necessary to be able to root out such correlated combinations. As usual, the penalty g is not unique and has to be set by the researcher. Analogous to Bai and Ng, in this paper we set

$$g = (m+1)\frac{\ln(\min\{N, \sqrt{T}\})}{\min\{N, \sqrt{T}\}}, \tag{24}$$

where the term $(m + 1)$ is there to account for the dimension of $V(\hat{F}(s))$. This ensures that $g \to 0$ and $\min\{N, \sqrt{T}\} \cdot g = O(\ln(\min\{N, \sqrt{T}\})) \to \infty$.

The estimated number of combinations can be difficult to interpret. This is where the Assumption 2 requirement that $r = (m + 1)k$ comes in. It ensures $(m+1)\hat{k}$ is consistent for r: that is,

$$P[(m+1)\hat{k} = r] \to 1, \tag{25}$$

which means that after scaling $\hat{k}$ can be interpreted as an estimator of r. Another reason for requiring that $r = (m + 1)k$ is that it simplifies the proofs. By using the same arguments as in Westerlund (2018), the results reported in this paper can be extended to the case when $r \leq (m + 1)k$. In this case, however, $\hat{r}$ is not necessarily consistent for r, but is instead consistent for the minimal number of combinations required to approximate the underlying factor structure (see Smeekes, 2015, for a similar discussion in the context of subpanel selection).

Remark 5. In order to appreciate the implications of Theorem 3 it is convenient to treat $\hat{\beta}$ as a function of $\hat{k}$. Let us therefore write $\hat{\beta}_P(\hat{k})$ for $\hat{\beta}_P$. Clearly,

$$\begin{aligned} P[\sqrt{NT}(\hat{\beta}_P(\hat{k}) - \beta) \leq \delta] &= P[\sqrt{NT}(\hat{\beta}_P(\hat{k}) - \beta) \leq \delta | \hat{k} = k]P(\hat{k} = k) \\ &\quad + P[\sqrt{NT}(\hat{\beta}_P(\hat{k}) - \beta) \leq \delta | \hat{k} \neq k]P(\hat{k} \neq k), \end{aligned}$$

where $\delta > 0$. Because $P(\hat{k} = k) \to 1$ and $P(\hat{k} \neq k) \to 0$ by Theorem 3, while the first term on the right-hand side converges to $P[\sqrt{NT}(\hat{\beta}_P(\hat{k}) - \beta) \leq \delta | \hat{k} = k] = P[\sqrt{NT}(\hat{\beta}_P(k) - \beta) \leq \delta]$, the second term converges to zero. It follows that

$$|P[\sqrt{NT}(\hat{\beta}_P(\hat{k}) - \beta) \leq \delta] - P[\sqrt{NT}(\hat{\beta}_P(k) - \beta) \leq \delta]| \to 0, \tag{26}$$

implying that Theorem 1 is unaffected by the estimation of k.

So far we have assumed that the candidates can be preordered, which is not very restrictive. In fact, as Donald and Newey (2001, p. 1156) point out in the classical IV context, "there will often be at least some information about which instruments are most important." Such information is, however, not necessary. If there is no natural ordering, then one possibility is simply to use an all-subset grid search, which is the approach used in the empirical application of Section 5. Grid search is feasible in applications where $k_{\max}$ is a relatively small number, which is likely to be the case in practice. For example, in our empirical illustration, $k_{\max} = 2$. If $k_{\max}$ is "large," then it is sometimes possible to group the candidates and to grid search among the groups (see, e.g., Donald & Newey, 2001). If this is not possible, one may follow, for example, Stock and Wright (2000) or Zheng and Loh (1995, 1997), and order the candidates according to an estimate of their correlation with C_i. This can be done by looking at $\hat{C}_i$, which we show in the supplementary Supporting Information to be consistent for the space spanned by C_i, provided that $N, T \to \infty$.[7] Alternatively, we may use $\bar{w}_i = T^{-1} \sum_{t=1}^{T} w_{i,t}$. The logic behind this latter approach is that $\bar{w}_i = T^{-1} \sum_{t=1}^{T} w_{i,t} = C_i' \overline{F} + \overline{u}_i = C_i' \overline{F} + o_p(1)$. This gives $m + 1$ correlations for each combination in z_i, which can be combined by taking, for example, the average of the absolute value of the correlations.

The idea of using the variation in C_i to say something about the nature of the cross-section dependence is not new but has been used before, although not in this particular context. Holly et al. (2010) used the CCE approach to infer the effect of income ($x_{i,t}$) on house prices ($y_{i,t}$) across US states. Motivated by the common practice in the capital asset pricing model (CAPM) literature, they regressed $y_{i,t} - x_{i,t}$ onto $\bar{y}_t - \overline{x}_t$ (and a constant), and used the estimated slopes as estimators of the factor loadings. They used the magnitude and sign of the estimated loadings to form roughly homogeneous groups, which were then characterized based on observables. Similar approaches have been used by Ludvigson and Ng (2007, 2009), for example, but then for studying estimated factors as opposed to estimated loadings.

5 | EMPIRICAL APPLICATION

Serlenga and Shin (2007) applied the CCEP estimator to a gravity equation of bilateral trade flows among 15 European countries (91 pairs) over the 1960–2001 period.[8] In their most general model, which they denoted "Case 3," they regressed real bilateral trade (TRADE) on the sum of the logged home and foreign country GDPs (TGDP), similarity in relative size (SIM), differences in relative factor endowments (RLF), the log real exchange rate (RER), a dummy variable which took on the value one when both countries belonged to the European community (CEE), and a dummy variable which took on the value one when both countries adopted a common currency (EMU). As common factor estimates they took the cross-section average of the observables plus the ECU/euro–US dollar real exchange rate (RERT), where the latter was added to capture the common influence of the USA.

In this section, we revisit the data of Serlenga and Shin (2007) and the model described in the previous paragraph. Our main reasons for doing so are to (i) illustrate the combination selection procedure and (ii) discuss the relevance of some of the most important assumptions and allowances in C^{3}E. As argued in Section 4, the process of finding suitable combinations is to a large extent reliant on pre-knowledge regarding the cross-section dependence. In particular, the combinations should be correlated with the loadings. Serlenga and Shin provided little or no motivation for their use of the CCEP estimator, or for why they considered common factors in the first place. They therefore did not offer any discussion of the nature of the cross-section dependence, although they did mention in their conclusions (see p. 377) that "it would be worth investigating the effect of globalization on transport costs more explicitly. For instance, transport and communication revolutions should lead to a dispersion of economic activity." Anderson and van Wincoop (2004)

[7]In the literature on generated regressors (see, e.g., Pagan, 1984), the sampling error induced by the generated regressors inflates the variance of the second-step regression estimator. This would be the case here too if N and /or T were fixed. However, in this study, N and T are both large, and under this assumption the estimated loadings are consistent. This means that asymptotically correlations based on estimated loadings are as good as correlations based on known loadings (see, e.g., Bai & Ng, 2006a, for a similar argument in the context of PC estimation).

[8]Serlenga and Shin (2007) also applied a Hausman–Taylor IV version of the CCEP estimator. However, as Baltagi (2010) pointed out, the results obtained by using this estimator seemed fragile.

TABLE 1 Correlation between estimated loadings and combinations

Variable	LAN ABS	LAN p value	BOR ABS	BOR p value	DIS ABS	DIS p value
TRADE	0.126	0.224	0.516	0.000	0.652	0.000
GDPT	0.076	0.465	0.274	0.007	0.282	0.005
RER	0.421	0.000	0.046	0.659	0.335	0.001
RLF	0.002	0.984	0.373	0.000	0.413	0.000
SIM	0.014	0.897	0.018	0.865	0.041	0.694
Average	0.128		0.245		0.345	

Note. The loadings are estimated by taking the time series average of the observables, as explained in the text. The variables therefore refer to the time series average associated with each variable. "ABS" refers to the absolute value of the estimated coefficients between the estimated loadings and each combination variable (LAN, BOR, and DIS).

introduced the concept of "multilateral resistance to trade," which refers to the fact that trade between two countries depends not only on the characteristics of the countries involved but also on the characteristics of other countries. Hence, as much of the recent empirical evidence shows (see, e.g., Herwartz & Weber, 2013), relative costs to trade matter. In a recent paper, Bertoli and Fernández-Huertas Moraga (2013) estimated a gravity equation to Spanish migration data. They showed how multilateral resistance to migration gave rise to cross-section dependence in the form of common factors. The authors also showed that the effect of the factors was likely to be heterogeneous across the cross-section, and that the extent of this heterogeneity depended in part on the cost of migration. Hence, in terms of the trade illustration considered here, the factor loadings are likely to be correlated with trade costs.

Motivated by the discussion of the previous paragraph, in this section we explore the possibility of using different measures of trade costs as means to combine the data. The by far the most common measure in the literature is physical distance (see Anderson & van Wincoop, 2004). In order to capture this, here we follow much of the previous literature (see again Anderson & van Wincoop, 2004, and the references provided therein), and use the distance between capital cities (DIS), and a dummy variable that takes the value one if the countries within a pair share a border (BOR). We also consider a vector of ones, and a dummy that is one if the countries have the same language (LAN), which is intended to capture in part similarity in cultural and historical backgrounds, in part an established network of translation (see, e.g., Melitz, 2008). To get a feeling for the correlation between each combination variable and the loadings, as discussed in Section 4, we check whether the combinations correlate with $\bar{w}_i$. The results are reported in Table 1. We see that while LAN is not very correlated, with a few exceptions, BOR and DIS are highly correlated with the elements of $\bar{w}_i$. We also see that the correlations reported for DIS are uniformly higher than those reported for BOR, and that they are highly significant (with one exception). Hence, based on these results, DIS stands out as the most preferred combination candidate. To decide on which combinations to use, we apply the IC criterion. We grid search over all possible combinations of the $k_{\max} = 4$ candidates. The lowest value of the IC is obtained when using DIS only with the vector of ones ending up in second place. Hence $\hat{k} = 1$, which is not that surprising given the relatively large number of regressors in this illustration.

The estimated factors are given by the DIS weighted cross-sectional averages of TRADE, GDPT, RER, RLF, and SIM. RERT is also included as well as a vector of ones, which is tantamount to the inclusion of country-pair-specific fixed effects. By construction, there is not much time variation in CEE and EMU. Therefore, in order to prevent multicollinearity with the vector of ones, the weighted averages of CEE and EMU are not included among the estimated factors (as in Serlenga & Shin, 2007). Figure 1 plots the estimated factors over time. The first thing to note is that, except possibly for the weighted average of SIM, all the series have marked trends.[9] This is very interesting because it is consistent with the results of Bun and Klaassen (2007), who found that the residuals from their fixed-effects model exhibited trending behavior over time. As a partial solution to the problem, they augmented their model with country-pair-specific linear time trends, which was shown to have a substantial effect on the results. In particular, the estimated GDP and currency union effects were markedly reduced when the trend was included—a finding that has received considerable attention in the subsequent literature. As a result, it is now quite common to allow not only for fixed effects but also for linear

[9]Because the factors are only identified up to a matrix rotation, we cannot interpret their level and sign. We can, however, interpret their behavior over time.

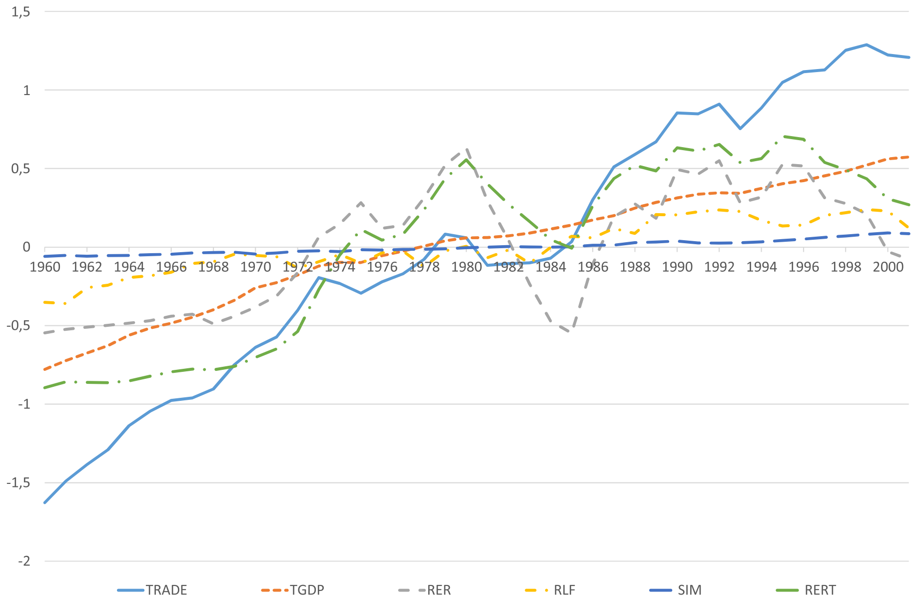

FIGURE 1 Plotting RERT and the DIS weighted cross-sectional averages of TRADE, GDPT, RER, RLF and SIM over time. For ease of comparison, all series have been normalized to have zero mean [Colour figure can be viewed at wileyonlinelibrary.com]

time trends. Of course, as Bun and Klaassen recognized themselves, such trends provide a very rough approximation to omitted trending variables, and, as already mentioned, the results are highly sensitive in this regard. This is important, because, as far as we are aware, no one has really tested the validity of the linear trend assumption, which means that we do not know if the sensitivity of the results stops here or if we have to consider more general models.

The generality of the common factor model considered in the present paper enables us to test the validity of the linear trend assumption. Note, in particular, that if the trend-augmented fixed effects model of Bun and Klaassen (2007) is correct, then it can be written in the same form as Equations 1 and 2, with $F_t = [1, \theta_t, t]'$ and $\lambda_i = [\alpha_i, 1, \delta_i]'$, such that $\lambda_i' F_t = \alpha_i + \theta_t + \delta_i \cdot t$. That is, the linear trend can be seen as a common factor, which fits in quite well with the trade cost story, as trade costs tend to decrease over time (see Bun & Klaassen, 2007). Hence, while there are some studies that allow for common factors (see, in particular, Cameraro et al., 2013; Serlenga & Shin, 2007), as already mentioned, their motivation has been generic, and no one has made the connection to the model of Bun and Klaassen.[10]

In order to investigate the validity of the linear trend assumption, we test whether the linear trend spans the same space as the true factors. This is done by applying the $A(j)$ statistic of Bai and Ng (2006b), which is the empirical rejection frequency of a sequence of point-wise two-sided α-level tests of equality between the spaces spanned by a linear trend and the estimated factors. If the spaces are the same, then $A(j)$ should be close to 2α. We set $\alpha = 0.05$, and obtain $A(j) \approx 0.548$, suggesting that a linear trend is not enough to capture the trending behavior of the factors.[11] Because the trend is not deterministic, we examine whether the trending behavior of the factors is stochastic. This is done by applying a version of the PANIC (PANICCA) approach (Bai & Ng, 2004; Reese & Westerlund, 2016). The idea is to estimate both the factors and idiosyncratic errors from the first-differenced data. The estimated components are then accumulated up to levels and subjected to a unit root test. The results are reported in Table 2. The estimated largest autoregressive root of the estimated factors all fall in the $[0.606, 0.908]$ range, and the augmented Dickey–Fuller (ADF) test does not provide any evidence against the unit root null. By contrast, the estimated idiosyncratic errors are all stationary at the 10% level or better. Hence, as might be expected, the trending behavior is not due to a simple linear time trend, but is rather due to a number of common stochastic trends. In order to determine how many distinct stochastic trends there are, we apply the MQ_c and MQ_F tests of Bai and Ng (2004), which both suggest that there are six stochastic trends: one per estimated factor (as in Cameraro et al., 2013, 2014).

The fact that F_t contains unit roots is inconsistent with Assumption 1(ii). As already pointed out, however, this assumption is not really necessary and can be relaxed without affecting the results. In fact, as Westerlund (2018) showed,

[10]It is also not clear if the PC-based approaches employed by Cameraro et al. (2013, 2014) are equipped to deal with the presence of trending factors.
[11]Bai and Ng (2006b) only consider principal components-based factor estimates. However, intuition suggests that the $A(j)$ statistic should work well also when applied to (weighted) cross-section averages, and unreported Monte Carlo results confirm this.

TABLE 2 Unit root test results for the estimated idiosyncratic errors and common factors

	Errors		Factors		
Variable	$P_{\hat{e}}^{c}$	***p* value**	**AR**	**ADF**	***p* value**
TRADE	5.011	0.000	0.708	-3.044	0.133
GDPT	5.051	0.000	0.908	-2.656	0.260
RER	1.506	0.066	0.739	-2.868	0.182
RLF	8.411	0.000	0.606	-2.990	0.147
SIM	4.409	0.000	0.814	-1.879	0.647
RERT			0.884	-1.738	0.715

Note. "Errors" and "Factors" refer to the estimated idiosyncratic errors and common factors, respectively. $P_{\hat{e}}^{c}$ is the combination of *p*-value test statistic of Bai and Ng (2004) based on the (weighted) cross-section averages instead of estimated principal component factors (as in Reese & Westerlund, 2016). "AR" refers to the estimated largest autoregressive root of each of the estimated factors with "ADF" being the associated unit root test statistic. The test regression for the idiosyncratic (common) unit root test is fitted with a constant (constant and trend) and the Schwartz Bayesian information criterion to select the lagged terms to include in order to control for error serial correlation. RERT does not vary across the cross-section and is treated as an observable common factor. Hence, for this variable, there is no idiosyncratic component.

except for some basic moment conditions, F_t is virtually unrestricted. Examples of permissible factors include polynomial time trends of any finite degree, seasonal and structural break dummies, factors with unknown heteroskedasticity over time, and factors of an unknown but finite order of integration. While his proof is for the CCEP estimator, the same arguments can be applied to C^3E, and the Monte Carlo results reported in the supplementary Supporting Information confirm this. The presence of unit root factors is therefore inconsequential for the results, which means that we can proceed as if Assumption 1(ii) was in fact met. The estimation results are reported in Table 3.[12] For comparison purposes, we also consider the very popular one- and two-way fixed-effects estimators and their trend-augmented counterparts.

According to Table 3 the fixed-effects results depend quite substantially on whether the linear time trend is included or not. In particular, except for RER in the one-way specification, the inclusion of the linear trend causes the estimated coefficients to drop. To take an extreme example, in the two-way specification, while the effect of EMU without trend is estimated to be 0.218, with the trend included it is −0.014, which is not even of the same sign. Hence, as argued by Bun and Klaassen (2007), proper accounting for the trend in TRADE is key. However, we also know that a linear trend is not enough, and that there is a need to account for common stochastic trends. In order to verify that this is indeed the case, the regression residuals are tested for unit roots using the four panel unit root tests of Smith et al. (2004), henceforth denoted by $\bar{t}$, $\overline{\mathrm{LM}}$, max, and min, which are based on combining individual ADF tests, and using a sieve bootstrap to account for weak cross-section (and serial) dependence in the errors. The fact that the dependence is not permitted to be strong is key, because it means that while the errors can be cross-correlated, they cannot contain unit root factors. In fact, the presence of unattended unit root factors will make the test biased towards accepting the unit root null. In agreement with this, we see that all four tests are unable to reject the unit root null, and that this is true even when the trend is included. Hence, again, a linear trend is not enough to account for the trending behavior of TRADE. We also computed the CD test discussed in Pesaran et al. (2008), which tests the null hypothesis of no cross-correlation in the residuals. According to the results, while the evidence against the null is reduced by the inclusion of time fixed effects, the null hypothesis is still rejected at all conventional significance levels. Hence the regression errors are correlated across country pairs, which is suggestive of omitted common factors.

Motivated by the above discussion we now go on to consider the C^3E results. The first thing to note is that, in contrast to before, now the CD test is insignificant and the hypothesis of a unit root in the regression errors is rejected. This is important for at least four reasons. The first reason is that, as discussed in Section 2, while the regression errors do not

[12]Both the C^3E estimator and its bias-corrected version (presented in the supplementary Supporting Information) were applied. However, since the results were identical down to the third decimal, which is partly expected because $N = 91 > T = 42$, in Table 3 we only report the results of the unadjusted estimator.

TABLE 3 Estimation results

	One-way FE				Two-way FE					
	No trend		Trend		No trend		Trend		C^{3}E	
	EST	*p* value	EST	*p* value	EST	*p* value	EST	*p* value	EST	*p* value
RER	0.061	0.000	0.211	0.000	0.084	0.000	0.071	0.000	0.056	0.000
GDPT	1.812	0.000	1.109	0.000	3.053	0.000	1.019	0.000	1.719	0.000
RLF	0.033	0.000	-0.002	0.658	0.018	0.000	-0.016	0.000	-0.011	0.152
SIM	1.172	0.000	0.778	0.000	1.422	0.000	0.743	0.000	1.111	0.000
CEE	0.309	0.000	0.252	0.000	0.319	0.000	0.203	0.000	0.050	0.000
EMU	0.085	0.000	-0.013	0.001	0.218	0.000	-0.014	0.002	0.000	0.919
CD	0.170	0.000	0.260	0.000	-0.009	0.000	-0.007	0.002	0.000	0.961
$\bar{t}$	5.068	1.000	-0.513	0.341	6.593	1.000	2.966	0.822	-18.331	0.000
$\overline{\text{LM}}$	-5.613	1.000	-2.706	0.357	-6.333	1.000	-5.528	0.872	17.807	0.000
max	1.886	1.000	-3.301	0.387	4.086	1.000	0.749	0.926	-21.258	0.000
min	-2.662	1.000	0.232	0.401	-3.365	1.000	-3.195	0.932	24.441	0.000

Note. "One-way FE" ("Two-way FE") refers to the model with cross-section (cross-section and time) fixed effects, which in the "Trend" case is augmented with cross-section-specific linear trends. The C^{3}E estimator is based on using DIS as a combination. The top panel reports the estimated slopes and their *p* values, and the bottom panel reports some tests of the estimated residuals. In particular, we report the CD test for the null hypothesis of no cross-correlation, and the $\bar{t}$, $\overline{\text{LM}}$, max and min unit root tests of Smith et al. (2004) that allows error cross-correlation. The *p* values associated with the unit root tests are based on bootstrapping using 499 bootstrap replications.

have to be cross-section independent, we do require that any correlation is of the weak form, which will be the case if they are uncorrelated. The results based on the CD test can therefore be taken as evidence in favor of the at-most-weak dependence assumption. The second reason is that by removing the cross-correlation from the errors the C^{3}E approach controls not only for error cross-section dependence per se, but also for endogeneity, as the factors are permitted to load on the regressors. Countries that trade a lot with each other may experience high economic growth, may be more likely to adopt a common currency, and so on. To the extent that reverse causality of this type is important, given the flexibility of the common factor component, C^{3}E estimator is expected to be more robust than the popular fixed-effects-based ordinary least squares estimator.[13] The third reason is that the regression errors have to be stationary for Assumption 1(i) to be met. The fourth and final reason has to do with the interpretation of the results when some of the factors are unit root nonstationary. Bun and Klaassen (2007) recognized that the variables of the gravity equation were likely to contain unit roots, and found evidence of cointegration using the test of Pedroni (2004; see Fidrmuc, 2009, for some confirmatory results using the same test). The problem is that this test is a so-called "first-generation" test that does not allow for cross-section dependence, and its properties become suspect when this condition is not met. The fact that the no cointegration null can be rejected when using this test is therefore likely to be due to size distortions. Our results show that the variables of the gravity equation are not cointegrated on their own, but that cointegration requires augmentation by F_t. That is, TRADE does not form a long-run relationship with its determinants unless the factors are included. This demonstrates quite clearly the importance of the common factors not only for estimation but also for interpretation purposes—a fact that seems to have gone largely unnoticed in the previous literature. Serlenga and Shin (2007) use the same data set as us. They are among the few that recognize the importance of allowing for common factors. The issue of nonstationarity and its implications for inference is, however, ignored.

Looking next at the C^{3}E-based estimation results we see that the effect of TGDP is positive and highly significant, even at the conservative 1% level, which is just as expected. The estimated effect, 1.719, is larger than in studies such as Cameraro et al. (2014) and Fidrmuc (2009), but is very close to the CCEP estimate reported by Serlenga and Shin (2007). The fact that the effects of SIM and CEE are positive and significant is in agreement with the trade cost story, and is therefore expected. The effect of RER is also positive. This means that a depreciation of the home currency leads to an increase in trade, which again is just as expected. RLF and EMU are insignificant, which means that trade is not affected by differences in factor endowments or by monetary union membership. While the effect of RLF is ambiguous (see Serlenga & Shin, 2007), the effect of EMU is expected to be positive. Rose and Stanley (2005) perform a meta-analysis of the results of 34 studies on the effect of currency unions on trade. According to their results, currency unions have a significant positive effect

[13]See Bun and Klaassen (2007) for some arguments for why endogeneity is unlikely to be very important and for some confirmatory empirical results.

that is estimated to lie between 0.26 and 0.64. Our EMU estimate is zero down to the third decimal, which is obviously much smaller than this. Our estimate is, however, consistent with those of Bun and Klaassen (2007) and Serlenga and Shin (2007), which account for the trend in trade. Looking across the estimators we see that the C^3E estimates are smaller than the corresponding fixed-effects estimates without trend, which is consistent with the finding of Bun and Klaassen (2007) that trend augmentation reduces the estimated GDP and currency union effects. However, we also see that the C^3E results differ quite markedly from the trend-augmented fixed-effects results, which reinforces the evidence already reported against the linear trend.

While we can of course speculate, which we have also done, the C^3E approach used here does not provide any explanation as to what the factors might represent. It does, however, enable us to control for the effects of the factors, which is enough if the purpose is to estimate the impacts of the already known determinants of trade. The results reported here suggest that there is a need for better understanding of the forces driving the common component of trade, and that credible theories are likely to include some kind of common factors.

6 | CONCLUSION

This paper considers the problem of consistent estimation of a factor-augmented panel regression model in which the number of factors, r, is potentially larger than the number of observables, $m + 1$. The estimator that we propose can be viewed as an extension of the CCEP estimator of Pesaran (2006), which is based on using the cross-section averages of the observables as proxies for the latent factors. While CCEP does allow $r > m + 1$, it does so at a cost. In particular, it is required that the factor loadings are independently distributed, which in most cases of practical relevance is likely to be violated. But even if the assumption is in fact satisfied, violations of $m + 1 \geq r$ are still costly. This is particularly true in the homogeneous slope case, in which a violation causes a reduction in the rate of consistency, from the usual $\sqrt{NT}$ rate to $\sqrt{N}$. In the present paper, we take this last feature of CCE as our starting point. The purpose is to provide a simple extension that preserves $\sqrt{NT}$ consistency without requiring independent loadings.

The idea behind the proposed C^3E approach is to use not only the cross-section average but also other (cross-section) combinations of the observables. By taking $k \geq 1$ such combinations we can allow $k(m + 1) \geq m + 1$ common factors without requiring independent loadings. In the analysis of the properties of the resulting pooled C^3E estimator we consider not only the standard scenario of homogeneous slopes but also the case when the slopes have a random distribution across the cross-section. We show that the estimator is consistent and asymptotically normal, with the rate of consistency depending on the heterogeneity of the slopes; if the slopes are homogeneous, the rate is the usual $\sqrt{NT}$, whereas if they are heterogeneous the rate is only $\sqrt{N}$. This is true if the combinations are known. If there is uncertainty over which combinations to use, an IC can be used to select the appropriate combinations.

ACKNOWLEDGMENTS

Previous versions of this paper were presented at the 2015 UvA-Econometrics Panel Data Workshop in Amsterdam, and the 9th Meeting of the Netherlands Econometric Study Group in Tilburg. The authors would like to thank workshop participants, and in particular Herman van Dijk (Co-Editor) and two anonymous referees for many valuable comments and suggestions. Westerlund thanks the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1. Westerlund and Karabiyik thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship.

REFERENCES

Anderson, J., & van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, *42*, 691–751.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*, 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*, 1229–1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191–221.

Bai, J., & Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, *72*, 1127–1177.

Bai, J., & Ng, S. (2006a). Confidence intervals for diffusion index forecasts and inference for factor augmented regressions. *Econometrica*, *74*, 1133–1150.

Bai, J., & Ng, S. (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, *113*, 507–537.

Bai, J., & Ng, S. (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory*, *26*, 1577–1606.

Baltagi, B. (2010). Narrow replication of Serlenga and Yongcheol (2007) gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, *25*, 505–506.

Baxter, M., & Kouparitsas, M. A. (2004). Determinants of business cycle comovement: A robust analysis. (*NBER working paper W10725*). Cambridge, MA: National Bureau of Economic Research.

Bertoli, S., & Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, *102*, 79–100.

Bun, M., & Klaassen, F. (2007). The euro effect on trade is not as large as commonly thought. *Oxford Bulletin of Economics and Statistics*, *69*, 473–496.

Cameraro, M., Gomez, E., & Tamarit, C. (2013). EMU and trade revisited: Long-run evidence using gravity equations. *The World Economy*, *36*, 1146–1164.

Cameraro, M., Gomez, E., & Tamarit, C. (2014). Is the "euro effect" on trade so small after all? New evidence using gravity equations with panel cointegration techniques. *Economics Letters*, *124*, 140–142.

Chudik, A., & Pesaran, M. H. (2015a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, *188*, 393–420.

Chudik, A., & Pesaran, M. H. (2015b). Large panel data models with cross-sectional dependence: A survey, In B. H. Baltagi (Ed.), *The Oxford Handbook on Panel Data*. (pp. 3–45). Oxford, UK: Oxford University Press.

Chudik, A., Pesaran, H. M., & Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal*, *14*, C45–C90.

Chudik, A., & Straub, R. (2011). Size, openness, and macroeconomic interdependence. (*Globalization and Monetary Policy Institute Working Paper 103*). Dallas, TX: Federal Reserve Bank.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, *92*, 1–45.

Conley, T., & Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, *17*, 303–327.

Dees, S., di Mauro, F., Pesaran, H. M., & Smith, V. L. (2007). Exploring the international linkages of the euro area: A global VAR analysis. *Journal of Applied Econometrics*, *22*, 1–38.

Donald, S. G., & Newey, W. K. (2001). Choosing the number of instruments. *Econometrica*, *69*, 1161–1191.

Eberhardt, M., Helmers, C., & Strauss, H. (2013). Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics*, *95*, 436–448.

Eberhardt, M., & Teal, F. (2011). Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, *25*, 109–155.

Fama, E., & French, K. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*, 55–84.

Fidrmuc, J. (2009). Gravity models in integrated panels. *Empirical Economics*, *37*, 435–446.

Forbes, K. J., & Chinn, M. D. (2004). A decomposition of global linkages in financial markets over time. *Review of Economics and Statistics*, *86*, 705–722.

Fracasso, A., & Marzetti, G. V. (2014). International R&D spillovers, absorptive capacity and relative backwardness: A panel smooth transition regression model. *International Economic Journal*, *28*, 137–160.

Fracasso, A., & Marzetti, G. V. (2015). International trade and R&D spillovers. *Journal of International Economics*, *96*, 138–149.

Gonçalves, S., & Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, *182*, 156–173.

Greenaway-McGrevy, R., Han, C., & Sul, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics*, *169*, 48–53.

Herwartz, H., & Weber, H. (2013). The role of cross-sectional heterogeneity for magnitude and timing of the euro's trade effect. *Journal of International Money and Finance*, *37*, 48–73.

Holly, S., Pesaran, H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, *158*, 160–173.

Holly, S., Pesaran, H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, *69*, 2–23.

Imbs, J. (2004). Trade, finance, specialization and synchronization. *Review of Economics and Statistics*, *84*, 723–734.

Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics*, *160*, 326–348.

Ludvigson, S., & Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, *83*, 171–222.

Ludvigson, S., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, *22*, 5027–5067.

Mastromarco, C., Serlenga, L., & Shin, Y. (2016). Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. *Journal of Applied Econometrics*, *31*, 281–297.

Melitz, J. (2008). Language and foreign trade. *European Economic Review*, *52*, 667–699.

Moon, R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, *83*, 1543–1579.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*, 777–787.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, *25*, 221–247.

Pedroni, P. (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the purchasing power parity hypothesis. *Econometric Theory*, *20*, 597–625.

Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with a multi-factor error structure. *Econometrica*, *74*, 967–1012.

Pesaran, H., & Chudik, A. (2013). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometrics Review*, *32*, 592–649.

Pesaran, H., Schuermann, T., & Weiner, S. (2004). Modelling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economic Statistics*, *22*, 129–162.

Pesaran, M. H., Smith, L. V., & Yamagata, T. (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics*, *175*, 94–115.

Pesaran, H. M., Ullah, A., & Yamagata, T. (2008). A bias-adjusted LM test of error cross section independence. *Econometrics Journal*, *11*, 105–127.

Phillips, P. C. B., & Hansen, B. (1990). Statistical inference in instrumental variables regression with I(1) variables. *Review of Economic Studies*, *57*, 99–125.

Reese, S., & Westerlund, J. (2016). PANICCA: PANIC on cross-section averages. *Journal of Applied Econometrics*, *31*, 961–981.

Reese, S., & Westerlund, J. (2018). Estimation of factor-augmented panel regressions with weakly influential factors. *Econometric Reviews*, *37*, 401–465.

Rose, A. K., & Stanley, T. D. (2005). A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys*, *19*, 347–365.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis*, *9*, 263–274.

Serlenga, L., & Shin, Y. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, *22*, 361–381.

Smeekes, S. (2015). Bootstrap sequential tests to determine the order of integration of individual units in a time series panel. *Journal of Time Series Analysis*, *36*, 398–415.

Smith, V., Leybourne, S., Kim, T.-H., & Newbold, P. (2004). More powerful panel unit root tests with an application to the mean reversion in real exchange rates. *Journal of Applied Econometrics*, *19*, 147–170.

Stock, J., & Wright, J. (2000). GMM with weak identification. *Econometrica*, *68*, 1055–1096.

Su, L., Jin, S., & Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*, *186*, 222–244.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Tutz, G., & Binder, H. (2007). Boosting ridge regression. *Computational Statistics and Data Analysis*, *51*, 6044–6059.

Westerlund, J. (2018). CCE in panels with general unknown factors. *Econometrics Journal*, *21*, 264–276.

Westerlund, J., & Urbain, J.-P. (2013). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters*, *119*, 247–250.

Westerlund, J., & Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, *185*, 372–377.

Zheng, X., & Loh, W.-Y. (1995). Consistent variable selection in linear models. *Journal of the American Statistical Association*, *90*, 151–156.

Zheng, X., & Loh, W.-Y. (1997). A consistent variable selection criterion for linear models with high-dimensional covariates. *Statistica Sinica*, *7*, 311–325.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Karabiyik H, Urbain J-P, Westerlund J. CCE estimation of factor-augmented regression models with more factors than observables. *J Appl Econ*. 2019;34:268–284. https://doi.org/10.1002/jae.2661